

컨텐츠 기반 P2P 파일 관리를 위한 분류 기법

A Classification Mechanism for Content-Based P2P File Manager

민 수 홍, 조 동 섭
(Suhong Min, DongSub Cho)

Abstract – P2P Systems have grown dramatically in recent years. Now many P2P systems have developed and been confronted by P2P technical challenges. We should consider how to efficiently locate desired resources. In this paper we integrated the existing pure P2P and hybrid P2P model. We try to keep roles of super peer in hybrid and concurrently use pure P2P model for searching resource. In order to improve the existing search mechanism, we present contents-based classification mechanism. Proposed system have the following features. This can forward only query to best peer using RI. Second, it is self-organization. A peer can reconfigure network that it can communicate directly with based on best peer. Third, peers can cluster each other through contents-based classification.

Key Words : Peer-to-Peer, best peer, self-configuration, classification

1. 서 론

최근 P2P 모델을 기반으로 한 시스템의 등장으로 다양한 자원을 효율적으로 이용할 수 있게 되었다. 기존의 클라이언트/서버의 경우, 모든 서비스가 서버에 집중되어 있어 인터넷의 자원인 정보, 대역폭, 컴퓨팅 자원을 활용하는 데 있어 한계점이 있다. 이러한 문제점을 해결하고자 P2P 모델이 등장하였다. P2P는 네트워크에 참여하는 모든 피어가 클라이언트와 서버의 역할을 동시에 수행한다. 이러한 특징 때문에 피어들은 분산되어 있는 자원을 자율적으로 공유할 수 있다. 따라서 P2P 환경에서 분산되어 있는 자원을 공유하는 문제는 중요한 이슘이며, 이에 대한 많은 연구가 현재 진행 중이다. 자원을 공유하는 문제는 피어의 자원을 어떻게 검색할 것인지의 문제이기도 하다. 피어의 자원에 대한 검색은 크게 3가지 P2P 모델 (순수 P2P, 중앙 집중형 P2P, 혼합형 P2P 모델)에 따라 나뉘어 질 수 있으며, 각 모델에 따라 세부적으로 나뉘어 진다. 본 논문에서 제안하는 P2P 시스템은 혼합형 P2P 모델과 순수 P2P 모델을 혼합한 모델로서 두 모델 장단점을 서로 보완하였다. 제안하는 P2P 구조는 피어와 수퍼 피어가 있는 혼합형 모델이며, 자원을 공유하는 방법으로 순수 P2P 모델의 flooding 방법을 이용하였다.

저자 소개

* 민 수 홍 : 이화여자대학교 컴퓨터학과 박사과정

** 조 동 섭 : 이화여자대학교 컴퓨터학과 정교수

그러나 기존의 순수 P2P 모델의 flooding 방식은 멀티캐스트 방식으로 네트워크 내에 있는 모든 피어에게 쿼리를 전송하여 대역폭 낭비 등을 초래 할 수 있으므로 본 논문에서는 기존의 RI 알고리즘을 수정하여 쿼리를 베스트 피어 중심으로 전송하며, 이에 대한 정보를 바탕으로 네트워크를 자율적으로 재구성하도록 하였다. 또한, 컨텐츠 중심으로 베스트 피어들을 클러스터링 하여 검색의 효율성을 높이고자 한다.

2. 관련 연구

현재 P2P 모델에 대한 많은 연구가 진행 중이다. 그 중 대표적인 P2P 모델은 서버가 존재하지 않는 순수한 P2P 모델과 디렉토리 서버를 둔 중앙 집중형 모델 그리고 이를 결합한 혼합형 모델로 나눌 수 있다. 그러나 초기에 연구 되어진 중앙 집중형 모델은 저작권 문제, 서버의 부하 문제 등으로 인해 연구의 진행이 미비한 상태이다. 따라서 이 장에서는 순수형 P2P 모델과 혼합형 P2P 모델에 대해 살펴보고자 한다.

2.1 순수 P2P (Pure P2P)

순수 P2P 모델은 네트워크 내에 공유되는 모든 자원의 인덱스 정보를 유지하는 중앙 서버가 존재하지 않는다. 따라서 피어는 스스로 다른 피어의 위치와 자원을 식별해야 한다. 피어가 자원을 찾기 위해서는 네트워크 상에 존재하는 피어들이 주기적으로 그들의 존재를 멀티캐스트를 이용해 상대 피어들에게 메시지를 보내 알려야 한다. 보통 메시지에는

TCP/IP 호스트 네임과 포트 번호가 포함된다. 이러한 정보를 통해 피어는 원하는 피어와 해당 자원들을 인접 피어에게 쿼리를 보내 요청할 수 있는데, 전형적인 쿼리 방법으로는 flooding을 들 수 있다. 피어가 멀티캐스트 방식으로 보낸 쿼리는 일정 반경 안에 있는 모든 피어들에게 전달된다. 메시지를 받은 피어들은 해당 자원이 있을 때 이를 요청한 피어에게 전달하게 된다. 이러한 검색 방법은 피어 간의 의존성이 낮기 때문에 참여, 탈퇴가 빈번한 P2P 환경에 적합한 구조이다. 그러나 피어를 찾기 위해 네트워크 상에 지나치게 많은 메시지를 발생시켜 대역폭을 낭비할 수 있으며, 피어를 관리하는 서버가 존재하지 않기 때문에 자원을 찾는데 어려움이 따른다는 단점이 있다. 대표적인 모델로 선 마이크로시스템의 JXTA, 그누텔라 등을 들 수 있다 [1].

2.2 혼합형 P2P (Hybrid P2P)

혼합형 P2P 모델은 중앙의 디렉토리 서버가 존재하지 않는 대신, 피어의 네트워크 참여, 탈퇴를 관리하는 서버 (또는 수퍼 피어)를 두고, 피어들을 관리하는 방법이다. 서버는 각 피어들간의 통신에 대한 중재 역할을 하는데, 연결되어 있는 피어들의 GUID (Global Unique Identifier), 공유 자원들에 대한 메타 데이터 등을 유지하며, 자원의 검색 기능과 기타 통신 기능 등을 제공한다. 혼합형 모델은 순수 P2P 모델과 비교했을 때 네트워크상에 있는 피어들을 쉽게 찾을 수 있으며 자원의 검색 또한 손쉽게 이루어진다는 장점이 있다. 반면, 이 모델의 경우 피어들은 여전히 서버의 의존도가 높으며, 피어의 수가 많아 질 경우 서버의 부하가 심해진다는 문제점이 있다. 대표적인 모델로는 Napster가 있다 [2, 4].

3. Peer-to-Peer System

본 장에서는 제안하는 시스템과 자원 공유를 위한 기법에 대해 설명하고자 한다.

3.1 전체 시스템 구조

본 논문에서 제안하는 P2P 시스템은 혼합형 P2P 시스템 구조와 순수 P2P 모델을 결합한 방법이다. 시스템 구조는 그림 1과 같다. 네트워크 타입은 피어와 수퍼 피어 두 가지 타입으로 구성된다. 수퍼 피어는 피어의 온라인/오프라인 판한 정보와 검색의 효율성을 위해 피어의 IP 주소를 관리한다. 그러나 기존의 혼합형 P2P 시스템은 수퍼 피어 (또는 중앙 서버)에 공유 정보가 집중되어 있기 때문에 많은 수의 피어가 자원을 검색 할 경우, 수퍼 피어에 부하를 줄 수 있으며, 수퍼 피어의 연결이 끊어졌을 경우 P2P 서비스를 수행할 수 없다. 또한 저작권 문제도 고려해야 한다. 따라서 본 논문에서는 수퍼 피어의 의존도를 줄이기 위하여 피어들은 선택적으로 순수 P2P 방식으로 서로 통신하고, 자원을 공유할 수 있도록 하였다. 본 논문에서 제안하는 시스템의 특징은 다음과 같다.

- 피어는 자원 검색을 위해 순수 P2P 방식과 혼합형 P2P 방식 중 선택하여 원하는 서비스를 제공받는다.

- 혼합형 P2P 방식을 선택할 경우, 피어는 수퍼 피어에서 제공하는 검색을 사용하며, 자원 공유는 피어 간의 다이렉트로 이루어진다.
- 순수 P2P 방식을 선택할 경우, 피어는 RI (Routing Indices) 알고리즘을 사용하여 피어의 쿼리를 전송한다. 이를 통해 인접 피어로 베스트 피어를 선별한다. 피어는 베스트 피어 정보를 바탕으로 네트워크를 재구성하여 자원 검색 시 응답 시간을 줄이며, 컨텐츠 별 클러스터링을 하여, 검색의 효율성을 높인다.

본 논문에서는 피어가 자원을 공유하기 위해서 두 가지 방법을 선택해서 사용할 수 있도록 하였다. 먼저, 피어는 기존의 혼합형 P2P 방법으로 컨텐츠 검색을 통해 요청하는 자원을 제공받을 수 있다. 이때 피어들의 자원을 랜덤하게 검색하여, 자원을 다이렉트로 제공받을 수 있다. 두 번째, 피어는 인접 피어들의 라우팅 인덱스 정보에 따라 인접해 있는 베스트 피어를 중심으로 쿼리를 전송할 수 있으며, 이를 이용해 네트워크를 재구성한다. 이 방법은 기존의 순수 P2P 방식인 멀티캐스트로 네트워크 내에 있는 모든 피어에게 쿼리를 flooding 하는 문제점을 개선할 수 있다. 또한 베스트 피어에게만 쿼리를 전송하여 랜덤하게 자원을 가진 피어들을 찾는 방법에 비해 응답 시간을 줄일 수 있다. 그리고 네트워크 재구성[5]을 통해 다이나믹한 P2P 환경을 반영할 수 있다. 따라서 베스트 피어를 중심으로 한 쿼리 전송을 이용하여 컨텐츠 별로 피어들을 그룹화 하여 클러스터링 하였으며 이를 통해 검색의 효율성을 높이고자 한다.

수퍼 피어는 피어들의 IP 주소와 피어의 네트워크 상태를 모니터링 한다. P2P 시스템의 경우, 피어의 네트워크 참여와 탈퇴가 빈번하게 이루어지는 점을 고려하여, 피어의 네트워크 참여 상태와 물리적인 상태 정보 (CPU, Storage 등)를 피어들에게 제공할 수 있도록 한다.

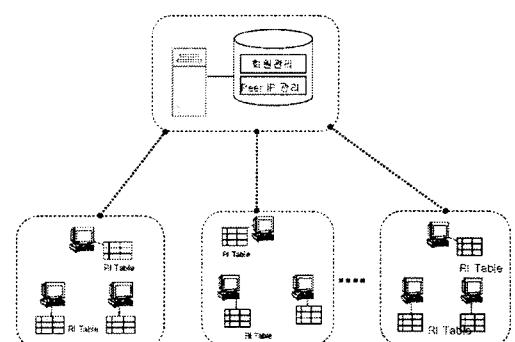


그림 2 전체 시스템 구조

3.2 자원 공유를 위한 분류 기법

피어는 자원 공유를 위해 두 가지 방법을 선택하여 사용한다. 먼저, 기존의 혼합형 시스템을 이용하여 랜덤하게 자원을 검색하는 방법이다. 이 방법은 요청하는 자원의 컨텐츠를 기반으로 랜덤하게 피어들의 공유 폴더를 검색한다. 각각의 피어들은 자원을 공유하기 위해 공유 폴더나 디렉토리를 지정한다.

두 번째 방법은 인접하는 베스트 피어에게 쿼리를 전송해

자원을 공유하는 방법이다. 인접 피어들 중 베스트 피어를 선별하기 위해 본 논문에서는 RI (Routing Indices) [3] 알고리즘을 이용하였다. RI는 인접 피어 중 베스트 피어로 선정하여 그에게 쿼리를 포워딩한다. 이는 자원을 요청한 피어에게 자원의 정확한 위치를 알려주는 방법이 아니라 피어가 요청한 자원과 관련된 자원을 많이 보유하고 있는 피어들의 경로를 알려 줌으로써 피어가 베스트 피어들을 중심으로 자원을 검색할 수 있도록 한다. RI를 이용한 검색 방법은 다음과 같다. 그림 2에서 보는 바와 같이 tree 형식으로 토플로지가 되어 있고, B, C, D 가 보유하고 있는 문서는 주제별로 그림 2의 테이블과 같다고 가정한다. 만약 A가 "database"와 "language" 관한 문서를 검색한다고 할 때, 피어 A는 인접한 피어 B, C, D의 문서의 수를 측정하며, (식 1)로 계산 할 수 있다.

$$\text{NumberOfDocuments} \times \prod_i \frac{\text{CRI}(si)}{\text{NumberOfDocuments}} \quad (\text{식 } 1)$$

따라서 B는 $\frac{20}{100} \times \frac{30}{100} \times 100 = 6$ 의 값을 가지며, C는 $\frac{0}{100} \times \frac{30}{100} \times 100 = 0$, D는 $\frac{100}{200} \times \frac{150}{200} \times 200 = 75$ 가 된다.

따라서 A의 베스트 피어는 D이며, A는 D에게 쿼리를 전송 한다 [3].

Path	# docs	Documents with topics:			
		DB	N	T	L
B	100	20	0	10	30
C	1000	0	300	0	50
D	200	100	0	100	150

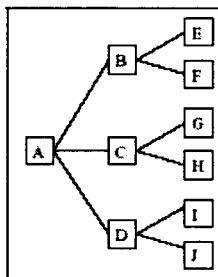


그림 3 Compound RI 알고리즘

본 논문에서 제안하는 시스템은 RI 기법을 이용해 인접 피어들 중 베스트 피어들을 선별하였으며, 이 정보를 바탕으로 네트워크를 자동으로 재구성할 수 있도록 한다. 본 논문에서는 기존의 RI 기법을 수정하여, RI를 통해 선정된 베스트 피어를 인접 피어로 다이렉트로 연결함으로써 그림 3과 같이 네트워크를 베스트 피어 중심으로 재구성하였다. 네트워크를 재구성할 경우, 피어들은 단순히 자원이 많은 피어를 베스트 피어로서 선정하는 것이 아니라 컨텐츠 별로 베스트 피어들을 인접 피어로 지정할 수 있다. 이를 통해 피어들의 자원은 컨텐츠 별로 분류되도록 한다. 네트워크를 재구성하는 방법은 다음과 같다. 먼저, 피어는 요청하는 자원을 라우팅 테이블 정보를 기반으로 베스트 피어에게 쿼리를 전송한다. 만약 쿼리 정보를 인접 베스트 피어가 리턴할 수 있다면, 자원에 대한 검색은 종료된다. 그렇지 않은 경우, 인접 피어는 자신의 베스트 피어에게 쿼리를 전송하게 되며, 쿼리를 받은 인접 피어의 베스트 피어는 자신의 공유 자원을 검색한다. 이를 통해 요청한 자원에 대해 인접 피어 중 베스트 피어만을 선별할 수 있다. 이와 같이 쿼리가 수행된 다음, 선별된 베스트 피어는 자동으로 다이렉트로 연결한다. 네트

워크 토플로지는 쿼리에 대한 흡수와 TTL 정보를 기반으로 해서 노드를 재구성한다. 이를 컨텐츠에 대해 적용했을 경우, 피어는 특별한 노력 없이 해당 컨텐츠에 대해 베스트 피어만으로 연결된 클러스터링에 자신의 피어를 추가시켜 자원을 공유할 수 있다. 예를 들어, 그림 3과 같이 피어 A가 관심을 갖는 컨텐츠를 음악과 영화라고 했을 때 피어는 라우팅 테이블 정보를 이용해 음악과 관련된 피어들과 영화에 관련된 피어들 중 베스트 피어만을 선별해 인접 피어들로 재구성한다.

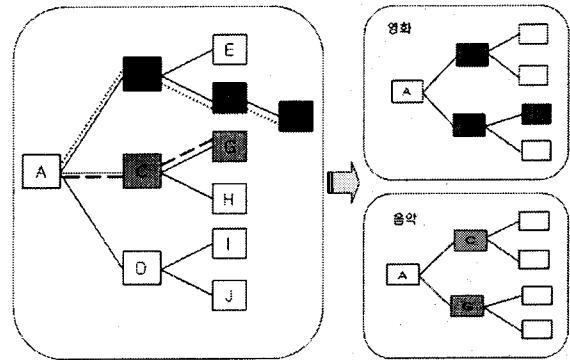


그림 4 컨텐츠별 분류

4. 결 론

본 논문에서는 기존의 순수형 P2P 모델과 혼합형 P2P 모델을 결합한 P2P 모델을 제안하였다. 제안한 P2P 모델은 혼합형 모델의 서비스 기능, 피어들의 상태 정보 등을 모니터링 기능을 유지하면서 동시에 다이나믹한 네트워크 환경에 적합한 순수형 P2P 모델을 적용해 수퍼 피어의 의존도를 줄이고자 하였다. 컨텐츠에 대한 검색은 RI 알고리즘을 수정하여 인접 피어들 중 베스트 피어만을 선택해 쿼리를 전송하였으며, 이를 통해 베스트 피어 중심으로 네트워크를 재구성하였다. 또한 자동으로 재구성된 베스트 피어는 컨텐츠 별로 피어들을 클러스터링 할 수 있다. 따라서 제안한 P2P 시스템은 수퍼 피어의 의존 없이 컨텐츠 별로 자원을 분류할 수 있으며, 이를 통해 검색의 효율성을 높일 수 있다.

참 고 문 헌

- [1] Andy Oram, "Peer-to-Peer" O'Reilly, March 2001.
- [2] B. Yang, H. Garcia-Molina, "Comparing hybrid peer-to-peers systems", In VLDB, 2001.
- [3] A. Crespo, H. Garcia-Molian, "Routing Indices For Peer-to-Peer Systems", Technical report, Computer Science Department, Stanford University, March 2002.
- [4] 김영진, 엄영익, "P2P 컴퓨팅 환경 기반의 디스커버리 기법", 한국정보과학회 정보과학회지, 제 22권 3호 178호, March 2004.
- [5] Wee Siong Ng, B. C. Ooi and K. L. Tan, "Best Peer: A Self-Configurable Peer-to-Peer System", International Conference on Data Engineering, March, 2002.